A Semi-Automated System to Annotate Communal Roosts in Large-Scale Weather Radar Data

Wenlong Zhao[1], Gustavo Perez[1], Zezhou Cheng[1],
Maria Carolina T. D. Belotti[2], Yuting Deng[2], Victoria F. Simons[2], Elske K. Tielens[3],
Jeffrey Kelly[3], Kyle G. Horton[2], Subhransu Maji[1], Daniel Sheldon[1]
[1]University of Massachusetts Amherst,[2]Colorado State University, [3]University of Oklahoma
`{wenlongzhao,gperezsarabi,smaji,sheldon}@cs.umass.edu`

Abstract

We have developed a semi-automated system to annotate communal roosts of birds and bats in weather radar data. This system comprises detection, tracking, confounder filtering, and human screening components. We have deployed this system to gather information on swallows from 612,786 scans taken from 12 radar stations around the Great Lakes over 21 years. The 15,628 annotated roost signatures have uncovered population trends and phenological shifts in swallows and martins. These species are rapidly declining aerial insectivores, and the data gathered has facilitated crucial sustainability analyses. While human screening is still required with the deployed system, we estimate that the screening process is approximately 7$\times$ faster than manual annotation. Furthermore, we found that incorporating temporal signals enhances the deployed detector's performance, increasing the mean average precision (mAP) from 48.7% to 56.3%. Our ongoing work aims to expand the analysis to bird and bat roosts at a continental scale.

1 Introduction

Weather radars around the globe continuously monitor the airspace and record activities of flying animals [1], offering opportunities to collect data about their behaviors at an unprecedented scale and resolution [2–6]. The NEXRAD radar network [7], for example, has archived more than 0.5 petabytes of data spanning 25 years and 159 radar stations that cover nearly the entire US [8]. Communal roosts of birds and bats can be studied with radar since they exhibit distinctive ring patterns when departing from roosting locations [9] (Fig. 2). Bird roosts in the US [10] are dominated by Purple Martins (*Progne subis*) or Tree Swallows (*Tachycineta bicolor*) [11–13], which are rapidly declining aerial insectivores in North America [14–16], and thus taxon-specific results about their ecology are of great interest. Bats offer significant pest suppression service to agricultural fields, but are at risk due to anthropogenic activities [17]. Past studies have employed manual annotation to identify roosts in radar [11, 18, 13, 12, 19]. Yet manual labeling at a large scale is costly and error-prone.

We develop a semi-automated system [20] for producing research-grade roost annotations from NEXRAD data (Figure 1). We construct a dataset with roost annotations to support model training. We develop open-sourced software for roost detection, tracking, and confounder filtering. Model predictions are imported to a web-based interface for screening. We deploy the system on 21 years of scans from 12 radar stations near the Great Lakes and report an estimated human annotation cost saving of $\sim 7\times$. The annotated data reveal that peak roosting activity time has advanced by 2.26 days per decade in the region potentially due to climate change [21]. We find that persistent roosts generally gather more birds, but most birds congregate in smaller roosts, and thus it is important to protect habitats of roosts of various sizes [22]. We enhance the detector to be able to process extra temporal scans by an adapter layer, which allows the model to recognize expanding movement

Computational Sustainability Workshop at NeurIPS 2023.

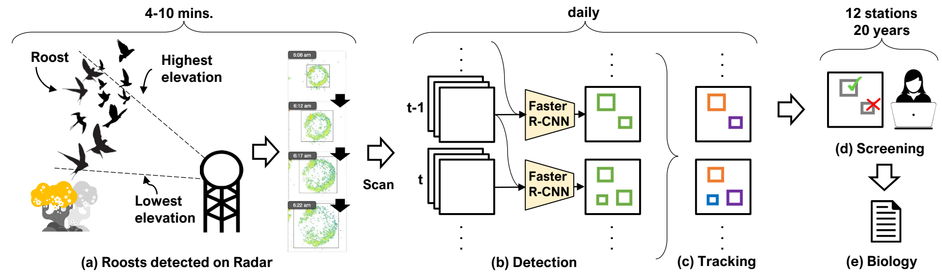


Figure 1: The proposed semi-automated system. (a) The exodus of birds and bats from roosting locations are visible as expanding rings in consecutive radar scans collected every 4 to 10 minutes. (b-c) Our system detects roosts in each scan and assembles those from consecutive scans in a day into tracks that represent roost signatures. (d) We develop a web-based interface to enable fast screening of roost signatures. (e) The screened data provide insights about population and phenology of the species.

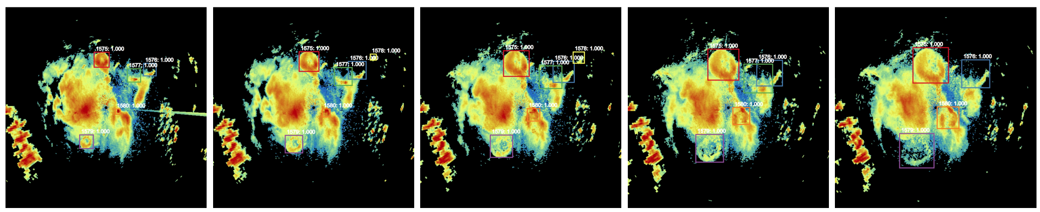


Figure 2: Training data examples: radar scans from the KDOX station on 10/1/2011 that are minutes apart. Swallow roosts exhibit ring-shaped patterns that expand over time.

dynamics of roosts and improves its performance from 48.7% to 56.3% mAP. We observe the system is effective on bat roosts without customization, and are extending the analyses to birds and bats at the continent scale.

2 Semi-Automated System for Roost Annotation in Weather Radar Data

Radar data preliminaries The NEXRAD radar network covers most of the U.S. territories and has archived radar data products since the 1990s. Each radar station performs a volume scan every 4-10 minutes by rotating the radar antenna around the vertical axis at several elevation angles to collect various products from cone-shaped slices of the surrounding airspace. At each elevation, each collected product is represented as a 2D array in polar coordinates indexed by range and azimuth (antenna direction in the horizontal plane). We render 600×600 Cartesian grids centered at the radar station from these 2D arrays by nearest neighbor interpolation [23] for $300\,\mathrm{km} \times 300\,\mathrm{km}$ regions.

Dataset We convert the roost dataset in [24] into the COCO object detection benchmark format, where radar scan lists and bounding box annotations are stored in json files. [24] chose data from 12 radar stations, KAMX, KDOX, KHGX, KJAX, KLCH, KLIX, KMLB, KMOB, KOKX, KRTX, KTBW, KTLH, and randomly sampled station-days to construct training, validation, and testing splits. Each station-day will have multiple radar scans that are minutes apart and swallow annotations for the scans. The annotations were initially manually labeled for ecological study [13]. After removing scans with rendering errors, we obtain 53 266 training, 11 599 validation, and 23 587 testing scans, with 37 619, 5139, and 10 942 single-scan roost labels. We apply scaling factors from [24] to normalize annotator style variation. Figure 2 illustrates the radar scans and annotations. The test set is for in-distribution evaluation, comprising data from the same station-years as the training data. In our deployment experiment described below, we verify the generalization of the trained detector by deploying it on unseen radar stations.

Roost detection, tracking, and filtering We train a Faster R-CNN detector with a ResNet101-FPN backbone and 45 anchors ranging from 16×16 to 512×512. The backbone is pretrained on ImageNet classification and MS-COCO detection. The detector takes three radar products as input channels: *reflectivity* at 0.5° and 1.5°, and *radial velocity* at 0.5°. Scaling the channels to 1200×1200 increases performance, likely because it is closer to the pretraining resolution. The deployed detector achieves 48.74 mean average precision (mAP) at IoU threshold 0.5 on our test set. We employ a greedy heuristic to assemble detections from consecutive scans into tracks [25], starting with high scoring detections and adding unmatched detections in neighboring scans with high overlap. We apply the Kalman filter [26] to smooth bounding box centers and radii in each roost track. We use dual-polarization radar products and external databases, when available, to remove rain and wind farm false positives and further reduce human screening efforts [27, 28]. More system details are reported in [20].

User interface and screening We build a web-based interface for visualization and human screening. Any track with ≥ 2 detections, ≥ 0.15 average detection score, and ≥ 1 detection with ≥ 0.5 score is labeled *roost*, considered "high confidence" roosts, and displayed with full opacity; other tracks are labeled *non-roost*, considered "low confidence" roosts, and displayed with low opacity. During screening, the labels can be updated. Clear roosts are labeled as *roost*; roosts contaminated by weather, anomalous propagation, and unknown noise are labeled as *weather-roost*, *ap-roost*, and *unknown-noise-roost*; duplicated tracks are labeled as *duplicate*; incomplete and drifting tracks are *bad-track*. If more than half of a scan is filled with weather or noise, the scan is abandoned.

Deployment To verify the general effectiveness of the system, we deployed it to annotate swallow roosts in scans from 12 radar stations in the Great Lakes region, including KAPX, KBUF, KCLE, KDLH, KDTX, KGRB, KGRR, KIWX, KLOT, KMKX, KMQT, and KTYX. These stations do not overlap with the training data stations. We downloaded scans from 30 minutes before to 90 minutes after local sunrise between June 1 and October 31 of 2000 to 2020. In the 120-minute window, we set 41 reference times spaced 3 minutes apart; for each time, we selected the scan closest to the reference time, if available, to be annotated. For predicted bounding boxes, we follow Chilson et al. [29] to convert reflectivity measurements to estimated bird counts in each roost.

Roost detection with temporal information Since roosts appear as distinctive expanding rings in consecutive radar scans when birds and bats depart from roosting locations, we further enhance the detector by leveraging channels from multiple temporal scans as inputs to capture the dynamics and potentially achieve better performance. We use a learnable adaptor [30] to map an arbitrary k-channel input array $x \in \mathbb{R}^{k \times n \times m}$ to a 3-channel image to be processed by the ResNet101-FPN backbone. More ablation study experiments are reported in [20].

3 Results and discussion

3.1 System deployment, improvement, and analysis

Great Lakes deployment statistics We successfully rendered $612\,786$ scans. Our algorithm predicted $31\,313$ "high-confidence" tracks assembled from $140\,036$ single-scan detections, and another $230\,088$ "low-confidence" tracks with $372\,594$ detections before screening. After screening, we identified $13\,860$ clean roost tracks, 477 that were contaminated by weather, 100 contaminated by anomalous propagation, and 1191 that contained unknown noise. These four categories together produced $15\,628$ roost tracks of $64\,620$ detections that can be useful in ornithology research.

Savings in human labeling efforts The screening in our deployment amounts to 183.6 annotator hours. We measured that annotating a station-day manually from scratch takes 2 minutes on average, and estimate that complete manual annotation of the Great Lakes station-years (153 days of 246 station-years with successfully rendered scans) will take 1254.6 annotator hours, or $6.83\times$ the screening time with our system.

Temporal information reduces false positives A significant technical advance was to incorporate temporal information from past scans to capture roost dynamics. When viewed in a single scan, weather and noise can have ring-shaped appearances as roosts do. However, weather usually moves in a straight trajectory, while roosts diverge from a point. We find that incorporating multiple scans improves the detector performance from 48.7% to 56.3% mAP at 0.5 IoU. Figure 3a shows an example where the model successfully discriminates roosts from weather.

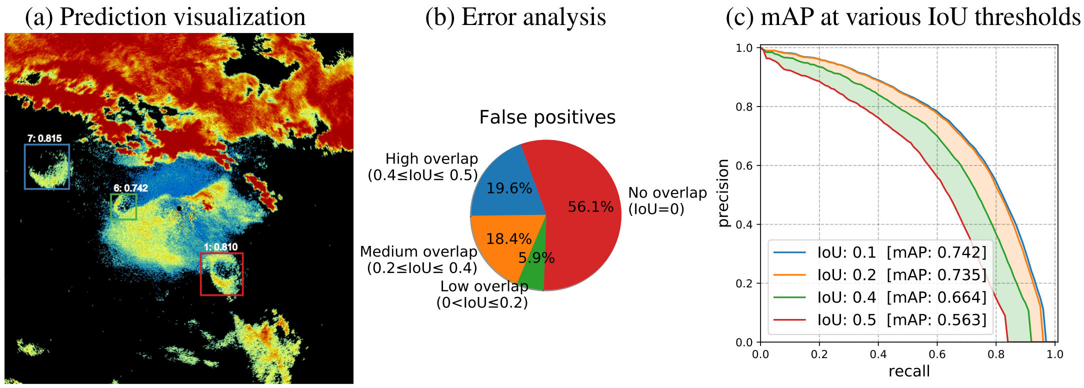


Figure 3: Temporal model analyses. **(a)** An example where the model successfully predicts three roosts and avoids weather confounders in a scan. **(b)** Among the false positives according to mAP at intersection-over-union (IoU) threshold 0.5, 56.1% are completely wrong predictions, while 44.9% are "near-misses". They partially overlap with ground truth and will be considered true positives according metrics with lower IoU thresholds. **(c)** Precision-recall curves at different IoU thresholds — higher reflects a stricter requirement. The mAP improves by 10.1% with IoU threshold 0.4 instead of 0.5, suggesting that a large number of detections are "near misses" that successfully find roosts but have bounding boxes with low overlaps compared to ground truth bounding boxes.

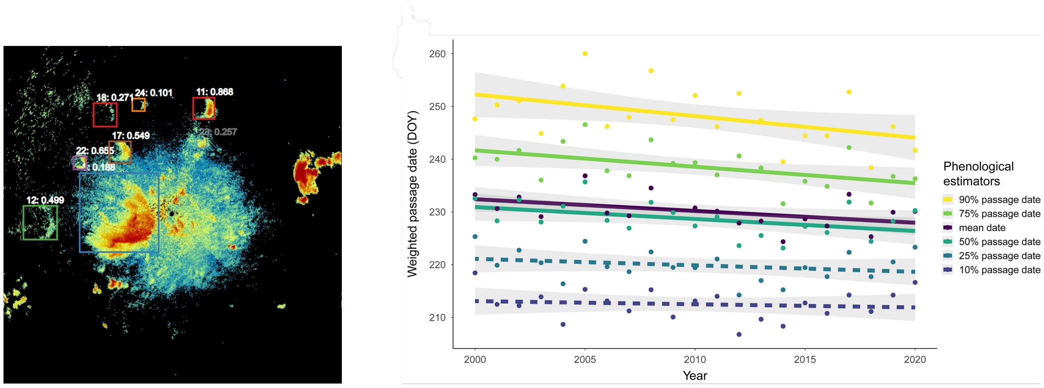
False positives are often near-misses Following the object detection literature in computer vision, we evaluate our detection models with the threshold of IoU> 0.5. Predicted bounding boxes with less than 0.5 IoU compared to any ground truth bounding box are considered false positives. However, we find that for the temporal model, 43.9% of these false positives actually partially overlap with some ground truth bounding boxes (Figure 3b). By reducing the IoU threshold from 0.5 to 0.4 in evaluation, we observe that the mean average precision (mAP) goes up significantly from 56.3% to 66.4% (Figure 3c). These indicate that the system has many "near-misses", which may be tolerable or adjusted during screening. A future version of the UI can allow users to adjust bounding box positions and sizes, although our current version mainly supports changing labels.

Deployment on bat roosts We deploy our temporal model on bat roosts around the KEWX radar station in Texas and observe that the system performs well at detecting and tracking bats without any customization. Figure 4(a) provides an example. The middle-left large radar return is at the Bracken Cave, the home to the world's largest bat colony [31]. Our ongoing work is further examining the system on bat roosts and aims to facilitate large-scale analyses of their behaviors.

3.2 Sustainability analysis

Phenological shift We aim to understand the interannual variations and long-term trend of swallows' roosting phenology at the Great Lakes. Passage dates were used as phenological estimators. n% passage date is the date when n% of the cumulative seasonal total of swallows are roosting in a region in a year. We use 10%, 25%, 50%, 75%, 90%, and mean dates. With several phenological estimators (different choices of n), we observe the trend of advancing passage dates over years at the Great Lakes (Figure 4b). In particular, the timing for peak and late roosting activity has become significantly earlier by 2-4 days per decade. Early roosting phases (10% and 25% dates). however, are not significantly advancing. This suggests that climate change may be impacting aerial insectivore phenology unevenly across different periods of their roosting stage, potentially leading to a shortened pre-migratory roosting season. This analysis provides the first quantification of the phenological trend in aerial insectivore roosting behavior and serves as a stepping stone for discovering mechanisms behind their phenological changes, which remain a priority for aerial insectivore conservation. We present more discussion in [21].

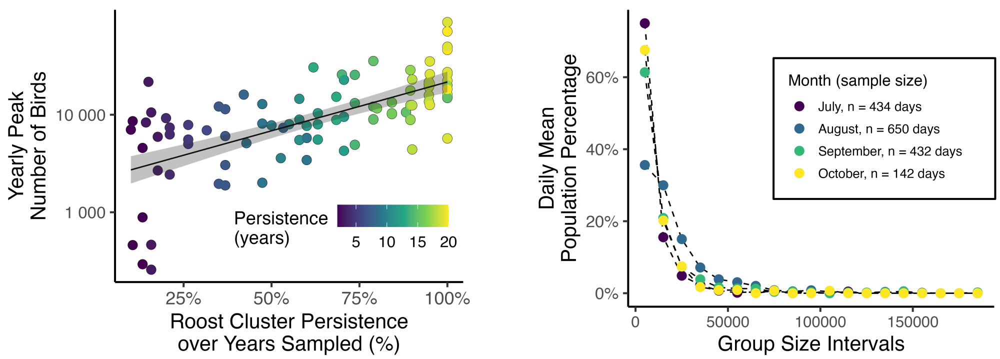
Roost persistence and size We set out to investigate how persistent the roosts were from 2000 to 2020 at the Great Lakes, and whether there is evidence that persistence is correlated with roost size. We have found that more persistent roost clusters gather more birds (Figure 5a), but that most



(a) System predicted bats on 6/30/2012 at KEWX around sunset. The system is trained on swallow annotations but can generalize to bat roosts.

(b) Each dot indicates the passage date of roosts in the Great Lakes region in a year. Fitted lines are derived from linear regression for each phenological estimator. Solid lines represent significant ($p < 0.05$) regression, and dotted lines non-significant. The passage dates advance across years, indicating phenological shifts.

Figure 4: Texas bat deployment results and Great Lakes swallow phenological analysis.



(a) Each dot is a roost signature found by mean shift clustering. There is a positive correlation between the peak roost size averaged over years and the percentage of years that the roost is detected among the years where radar data are available for the station (persistence). Colors indicate the number of years when a roost is recorded by radar data.

(b) Percentage of the entire Great Lakes swallow population that shows up in roosts of various sizes binned by 10,000 individuals. The population of each roost is computed per day and averaged over each month. We observe that a majority of swallow population gathers in roosts of less than 20,000 individuals.

Figure 5: Persistence and size analyses for swallow roosts around the Great Lakes.

bird individuals congregate in roosts of smaller sizes (Figure 5b). This suggests that protecting large predictable roosts alone may not be sufficient to safeguard most individuals of swallows and martins during their roosting stage and revert the recent rapid declines observed in these species. We present more details in [22].

4 Conclusion

We present a semi-automated system to annotate bird and bat roosts in weather radar data. In our deployment experiment on 21 years of radar data at 12 radar stations around the Great Lakes, the system reduces human labeling costs by 6.83× and the collected swallow annotations enable critical sustainability analyses about their population trends and phenological shifts. We find that the detector can be significantly improved by incorporating temporal information, since roosts exhibit expanding dynamics. The work represents a collaboration between computer scientists and ecologists to develop tools to rapidly enable radar data analysis at scale.

References

[1] Thomas H. Kunz, Sidney A. Gauthreaux, Jr., Nickolay I. Hristov, Jason W. Horn, Gareth Jones, Elisabeth K.V. Kalko, Ronald P. Larkin, Gary F. McCracken, Sharon M. Swartz, Robert B. Srygley, et al. Aeroecology: probing and modeling the aerosphere. *Integrative and comparative biology*, 48(1):1–11, 2008.

[2] Sidney A. Gauthreaux, Jr. Weather radar quantification of bird migration. *BioScience*, 20(1):17–19, 1970.

[3] Bruno Bruderer. The study of bird migration by radar part 1: The technical basis. *Naturwissenschaften*, 84 (1):1–8, 1997.

[4] Sidney A. Gauthreaux, Jr. and Carroll G. Belser. Displays of bird movements on the WSR-88D: patterns and quantification. *Weather and Forecasting*, 13(2):453–464, 1998.

[5] Sidney A. Gauthreaux, Jr., Carroll G. Belser, and Donald Van Blaricom. Using a network of WSR-88D weather surveillance radars to define patterns of bird migration at large spatial scales. In *Avian migration*, pages 335–346. Springer, 2003.

[6] Adriaan M. Dokter, Felix Liechti, Herbert Stark, Laurent Delobbe, Pierre Tabary, and Iwan Holleman. Bird migration flight altitudes studied by a network of operational weather radars. *Journal of The Royal Society Interface*, 8(54):30–43, 2011. doi: 10.1098/rsif.2010.0116. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2010.0116`.

[7] Timothy D. Crum and Ron L. Alberty. The WSR-88D and the WSR-88D operational support facility. *Bulletin of the American Meteorological Society*, 74(9):1669–1687, 1993.

[8] Steve Ansari, Stephen Del Greco, Edward Kearns, Otis Brown, Scott Wilkins, Mohan Ramamurthy, Jeff Weber, Ryan May, Jed Sundwall, Jeff Layton, Ariel Gold, Adam Pasch, and Valliappa Lakshmanan. Unlocking the potential of NEXRAD data through NOAA's big data partnership. *Bulletin of the American Meteorological Society*, 99(1):189–204, 2018. doi: 10.1175/BAMS-D-16-0021.1. URL `https://doi.org/10.1175/BAMS-D-16-0021.1`.

[9] Kevin R Russell and Sidney A Gauthreaux Jr. Use of weather radar to characterize movements of roosting purple martins. *Wildlife Society Bulletin*, pages 5–16, 1998.

[10] Kevin R Russell, David S Mizrahi, and Sidney A Gauthreaux Jr. Large-scale mapping of purple martin pre-migratory roosts using wsr-88d weather surveillance radar. *Journal of Field Ornithology*, pages 316–325, 1998.

[11] David W. Winkler. Roosts and migrations of swallows. *Hornero*, 21(2):85–97, 2006.

[12] Eli S. Bridge, Sandra M. Pletschet, Todd Fagin, Phillip B. Chilson, Kyle G. Horton, Kyle R. Broadfoot, and Jeffrey F. Kelly. Persistence and habitat associations of Purple Martin roosts quantified via weather surveillance radar. *Landscape Ecology*, 31(1):43–53, 2016.

[13] Andrew J. Laughlin, Daniel R. Sheldon, David W. Winkler, and Caz M. Taylor. Quantifying non-breeding season occupancy patterns and the timing and drivers of autumn migration for a migratory songbird using Doppler radar. *Ecography*, 39(10):1017–1024, 10 2016. ISSN 1600-0587. doi: 10.1111/ecog.01988. URL `http:https://doi.org/10.1111/ecog.01988`.

[14] Silke Nebel, Alex Mills, Jon D. McCracken, and Philip D. Taylor. Declines of aerial insectivores in North America follow a geographic gradient. *Avian Conservation and Ecology*, 2010.

[15] Kevin C. Fraser, Bridget J. M. Stutchbury, Cassandra Silverio, Patrick M. Kramer, John Barrow, David Newstead, Nanette Mickle, Bruce F. Cousens, J. Charlene Lee, Danielle M. Morrison, Tim Shaheen, Paul Mammenga, Kelly Applegate, and John Tautin. Continent-wide tracking to determine migratory connectivity and tropical habitat associations of a declining aerial insectivore. *Proceedings of the Royal Society of London B: Biological Sciences*, 2012. ISSN 0962-8452. doi: 10.1098/rspb.2012.2207. URL `http://rspb.royalsocietypublishing.org/content/early/2012/10/22/rspb.2012.2207`.

[16] Kenneth V. Rosenberg, Adriaan M. Dokter, Peter J. Blancher, John R. Sauer, Adam C. Smith, Paul A. Smith, Jessica C. Stanton, Arvind Panjabi, Laura Helft, Michael Parr, and Peter P. Marra. Decline of the north american avifauna. *Science*, 366(6461):120–124, 2019. doi: 10.1126/science.aaw1313.

[17] Ruscena Wiederholt, Laura López-Hoffman, Colleen Svancara, Gary McCracken, Wayne Thogmartin, Jay E Diffendorfer, Brady Mattson, Kenneth Bagstad, Paul Cryan, Amy Russell, et al. Optimizing conservation strategies for mexican free-tailed bats: a population viability and ecosystem services approach. *Biodiversity and Conservation*, 24:63–82, 2015.

[18] Andrew J. Laughlin, Caz M. Taylor, David W. Bradley, Dayna Leclair, Robert G. Clark, Russell D. Dawson, Peter O. Dunn, Andrew Horn, Marty Leonard, Daniel Sheldon, Dave Shutler, Linda A. Whittingham, David W. Winkler, and D. Ryan Norris. Integrating information from geolocators, weather radar and citizen science to uncover a key stopover area for an aerial insectivore. *The Auk*, 130(2):230–239, 2013.

[19] Jeffrey F. Kelly and Sandra M. Pletschet. Accuracy of swallow roost locations assigned using weather surveillance radar. *Remote Sensing in Ecology and Conservation*, 2017.

[20] Gustavo Perez, Wenlong Zhao, Zezhou Cheng, Maria Carolina TD Belotti, Yuting Deng, Victoria F Simons, Elske Tielens, Jeffrey F Kelly, Kyle G Horton, Subhransu Maji, et al. Using spatio-temporal information in weather radar data to detect and track communal bird roosts. *bioRxiv*, pages 2022–10, 2022.

[21] Yuting Deng, Maria Carolina TD Belotti, Wenlong Zhao, Zezhou Cheng, Gustavo Perez, Elske Tielens, Victoria F Simons, Daniel R Sheldon, Subhransu Maji, Jeffrey F Kelly, et al. Quantifying long-term phenological patterns of aerial insectivores roosting in the great lakes region using weather surveillance radar. *Global Change Biology*, 29(5):1407–1419, 2023.

[22] Maria Carolina TD Belotti, Yuting Deng, Wenlong Zhao, Victoria F Simons, Zezhou Cheng, Gustavo Perez, Elske Tielens, Subhransu Maji, Daniel Sheldon, Jeffrey F Kelly, et al. Long-term analysis of persistence and size of swallow and martin roosts in the us great lakes. *Remote Sensing in Ecology and Conservation*, 2023.

[23] J. Anthony Parker, Robert V. Kenyon, and Donald E. Troxel. Comparison of interpolating methods for image resampling. *IEEE Transactions on medical imaging*, 2(1):31–39, 1983.

[24] Z. Cheng, S. Gabriel, P. Bhambhani, D. Sheldon, S. Maji, A. V, and D. Winkler. Detecting and tracking communal bird roosts in weather radar data. *Association for the Advancement of Artificial Intelligence (AAAI)*, 2020.

[25] X. Ren. Finding people in archive films through tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR*, 2008.

[26] Rudolf Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960. ISSN 0021-9223. doi: 10.1115/1.3662552. URL `https://doi.org/10.1115/1.3662552`.

[27] Adriaan M. Dokter, Peter Desmet, Jurriaan H. Spaaks, Stijn van Hoey, Lourens Veen, Liesbeth Verlinden, Cecilia Nilsson, Günther Haase, Hidde Leijnse, Andrew Farnsworth, Willem Bouten, and Judy Shamoun-Baranes. bioRad: biological analysis and visualization of weather radar data. *Ecography*, 2018.

[28] B Hoen, J Diffendorfer, J Rand, L Kramer, C Garrity, and H Hunt. United states wind turbine database. us geological survey, american wind energy association, and lawrence berkeley national laboratory data release: Uswtdb v2. 1. *United States Wind Turbine Database*, 2019.

[29] Phillip B. Chilson, Winifred F. Frick, Phillip M. Stepanian, J. Ryan Shipley, Thomas H. Kunz, and Jeffrey F. Kelly. Estimating animal densities in the aerosphere using weather radar: To Z or not to Z? *Ecosphere*, 3 (8), 2012. doi: 10.1890/ES12-00027.1.

[30] G. Perez and S. Maji. Domain Adaptors for Hyperspectral Images. *26TH International Conference on Pattern Recognition (ICPR), 2022*, 2022.

[31] Rodrigo A Medellin, Ruscena Wiederholt, and Laura Lopez-Hoffman. Conservation relevance of bat caves for biodiversity and ecosystem services. *Biological conservation*, 211:45–50, 2017.